ABSTRACT
We introduce and evaluate a block-iterative Fisher scoring (BFS) algorithm for emission tomography. Regularization is achieved by penalized likelihood with a general quadratic penalty. When the algorithm converges, it converges to the unconstrained maximum penalized likelihood (MPL) solution.

In a simulated data set, constrained BFS achieves a higher penalized likelihood in fewer iterations than other block-iterative algorithms which are designed for non-negatively constrained penalized reconstruction.

Key Words: Block-iterative Fisher scoring, emission tomography, OS-EM, BSREM, OS-SPS.

1. INTRODUCTION
Fisher scoring is an efficient, stable statistical algorithm for parameter estimation. Reviews of its application may be found in [1]. This approach requires the solution of a linear system of equations involving Fisher’s information matrix in each iteration.

In tomography, Fisher scoring (or its modification) is competitive with Shepp and Vardi’s reconstruction based on the Expectation-Maximization (EM) algorithm [2]. The efficiency of the Fisher scoring method in the context of maximum likelihood (ML) is evidenced by the work of Tak [3], where it is shown that two to three iterations of Fisher scoring may provide sufficient numerical accuracy. It is, however, impractical to implement Fisher scoring directly in emission and transmission tomography.

We will develop block-iterative reconstruction methods maximizing penalized likelihood functions with a quadratic penalty. These algorithms implement block-iterative schemes within each Fisher scoring iteration.


The rest of the paper is arranged as follows. Section 2 defines Fisher scoring procedure for MPL estimation. Section 3 explains how to implement block-iterative Fisher scoring schemes to obtain MPL estimates with a quadratic penalty. Simulation results are reported in Section 4 and concluding remarks are given in Section 5.

2. FISHER SCORING FOR EMISSION TOMOGRAPHY
In what follows, \( x_j \geq 0 \) denotes the expected emission activity of pixel \( j, j = 1, \ldots, p \), and \( x \) denotes the \( p \times 1 \) vector of all \( x_j \). Besides, we let \( y_i \geq 0, i = 1, \ldots, n \), be the count on the \( i \)-th bin of the camera and \( y \) be the vector of all \( y_i \). Assume \( y_i \) are independent Poisson random variables. Following, for example, Ahn and Fessler [8], expected value \( \mu = E(y|x) \) is linked to the unknown \( x \) by

\[
\mu = Ax + \eta,
\]

where \( \eta \) is the mean background events which is assumed known and the elements \( a_{ij} \) of \( A \) represent the conditional probability of a photon being detected by camera bin \( i \) given it was emitted from pixel \( j \). We assume \( \sum_{j=1}^{p} a_{ij} > 0 \) for any given \( i \).

In the following derivation we shall assume all \( \mu_i > 0 \). This condition is automatically satisfied when all components of \( x \) are non-negative and \( \eta > 0 \). We comment further on this assumption for the case of \( \eta = 0 \) at the end of this section.

It is well known that estimating the emission activity vector \( x \) from observations \( y_i \) is ill-conditioned; the estimate must be regularized. We adopt the MPL approach, i.e., estimate \( x \) by maximizing the penalized log-likelihood \( l(x) \):

\[
l(x) = \sum_{i=1}^{n} \{-\mu_i + y_i \log \mu_i\} - \frac{1}{2} h J(x),
\]

for \( x \in D \), where \( D = \{ x : \mu > 0 \} \). In (2), \( h > 0 \) is the smoothing parameter and \( J(x) \) the penalty function. The
block-iterative method introduced in this paper is defined for quadratic penalties $J(x) = x^T \tilde{R} x$, where $\tilde{R}$ is assumed symmetric and positive definite. We absorb $h$ by defining $R = h\tilde{R}$ hereafter, referring to $R$ as the penalty matrix.

Let $l'(x^{(k)})$ and $l''(x^{(k)})$ represent the first and second derivatives (with respect to $x$) of $l(x)$ evaluated at $x^{(k)}$, within the domain of $l$. At the $(k + 1)$-th iteration the Fisher scoring algorithm provides the update $x^{(k+1)}$:

$$x^{(k+1)} = x^{(k)} + [F^{(k)}]^{-1} l'(x^{(k)}),$$

where $F^{(k)} = F = E(-l''(x))$ evaluated at $x^{(k)}$. We comment that the Fisher scoring algorithm may crash if it returns, at any iteration $k$, $\mu^{(k)} < 0$. In this case $F$ may not be positive definite. Let $A^T$ be the transpose of $A$ and $V(\mu)$ be a diagonal matrix with diagonal $\mu_1, \ldots, \mu_n$. For emission tomography, $x^{(k+1)}$ of (3) can be obtained by solving

$$(A^T W(k) A + R) x^{(k+1)} = A^T W(k) z.$$  

where $W(k)$ is the inverse of $V(\mu^{(k)})$ and $z = (y - \eta)$.

Equation (4) can be expressed in a dual form which then can be solved block-iteratively by partitioning $y$. The dual form makes use of the following matrix identity. Assume $V$ and $R$ are both non-singular; then, setting $W = V^{-1}$, we have

$$(A^T W A + R)^{-1} A^T W = R^{-1} A^T (AR^{-1} A^T + V)^{-1}.$$  

This leads to a procedure that solves (4) in two stages: firstly, solve the linear system

$$(AR^{-1} A^T + V(\mu)) \xi^{(k+1)} = z$$

in introduced variable $\xi^{(k+1)}$; then secondly, backproject $\xi^{(k+1)}$ and filter the resulting image to update $x$:

$$x^{(k+1)} = R^{-1} A^T \xi^{(k+1)}.$$  

Note that $\xi^{(k+1)}$ can be interpreted as a weighted residual: $\xi^{(k+1)} = W(k) (z - A x^{(k+1)})$.

This algorithm departs from the Fisher scoring algorithm by substituting an inexact solution for $\xi^{(k+1)}$ in equation (6). We still use $\xi^{(k+1)}$ henceforth to denote the inexact solution of equation (6). This inexact solution can be obtained block-iteratively. From $\xi^{(k+1)}$, we obtain $x^{(k+1)}$ by equation (7).

Because of the assumption that $R$ is invertible, this Fisher scoring algorithm cannot be applied for ML reconstruction. However, a simple way around the problem is that a very small $h$ (such as $h = 10^{-12}$) will give a good approximation to ML solution.

If, during iterations, it happens that $\mu_i^{(k)} \leq \epsilon$, where $\epsilon$ is a small threshold, we substitute $\epsilon$ for $\mu_i^{(k)}$ in the next iteration.

Ahn and Fessler [8] approach the possibility of $\mu_i = 0$ by modifying $l(x)$ such that it is well-behaved on $\mu_i = 0$. They replace $l(x)$ by a quadratic approximation when $\mu_i < \epsilon$. The modified objective function has the same maximizer(s) as the original.

3. BLOCK-ITERATIVE FISHER SCORING METHODS

3.1. Algorithms

We use a block SOR iterative splitting scheme [9, Chapter 14] to define sub-iterations that provide an approximate solution to equation (6).

Let $\{C_m\}_{m=1}^{M}$ be exclusive and exhaustive subsets of the index set $\{1, \ldots, n\}$. $C_m$ provides the index set for the $m$-th block. Vectors $z, \mu$ and matrix $A$ are partitioned into $M$ corresponding blocks: $z = (z_1, \ldots, z_M)^T$, $\mu = (\mu_1, \ldots, \mu_M)^T$ and $A = (A_1^T, \ldots, A_M^T)^T$. Similarly the diagonal matrices $V$ and $W$ are partitioned in diagonal blocks $V_1, \ldots, V_M$ and $W_1, \ldots, W_M$ respectively.

In performing sub-iterations we can either run a single-pass through all subsets or run multiple passes. The advantage of a multi-pass is that it can save computations due to less updating of $W$ while achieving similar results.

Let $b$ be number of sub-iterations (or passes). Let $r$ index sub-iterations, $r = 0, \ldots, b-1$. After the $r$-th sub-iteration the estimate of $\xi$ is denoted by $\xi^{(k+1)}_r$.

Substituting $A$ and $V$ partitions into (6) provides a partition to $AR^{-1}A^T + V(k)$; its diagonal blocks are $A_m R^{-1} A_m^T + V_m(k)$ and off-diagonal blocks are $A_m R^{-1} A_t^T$. Vector $\xi(k)$ is also partitioned with its $m$-th block denoted by $\xi_m^{(k)}$. The linear equations in (6) are divided into $M$ sub-equations:

$$z_m - A_m R^{-1} \sum_{t=1}^{M} A_t^T \xi_t^{(k+1)} = V_m(k) \xi_m^{(k+1)} - V_m \xi_m^{(k+1)} = 0,$$

$m = 1, \ldots, M$. Note matrix $V(k)$ remains unchanged in sub-iterations.

Let $\{Q_m^{(k)}\}_{m=1}^{M}$ be a sequence of non-singular matrices (usually dependant on $x^{(k)}$) with dimensions determined by the block sizes, then, by $\{Q_m^{(k)}\}$ and the SOR scheme, $\xi^{(k+1)}$ in (6) is updated block-wisely by

$$\xi_m^{(k+1)} = \xi_m^{(k+1)} + \omega Q_m^{(k)} \xi_m^{(k+1)} - \epsilon_m^{(k+1)}$$

where $\omega$ is a relaxation parameter, and $\epsilon_m^{(k+1)}$ is:

$$\epsilon_m^{(k+1)} = (z_m - A_m R^{-1} \gamma^{(m-1)} - V_m \xi_m^{(k+1)}),$$

where $\gamma^{(m-1)} = \sum_{t=1}^{m-1} A_t^T \xi_t^{(k+1)} + \sum_{t=m}^{M} A_t^T \xi_t^{(k+1)}$. Note that $\gamma^{(m)}$ can be cheaply computed using $\xi$:

$$\gamma^{(m)} = \gamma^{(m-1)} + A_m^T \xi_m^{(m-1)} - \epsilon_m^{(m-1)}.$$  

These considerations lead to a general BFS procedure summarized below.

**Block Fisher Scoring Algorithm (BFS)**

**Initialize:** Set $k \leftarrow 0$. Choose $\xi^{(0)}$ and $x^{(0)} \geq 0$ to satisfy $x^{(0)} = R^{-1} A^T \xi^{(0)}$.  

---

[154]
Sub-iterations:
1. Initialize inner loop: Set initial values \( r \leftarrow 0 \), \( \xi_{k,0} \leftarrow \xi^{(k)} \), \( \tilde{x}^{(0)} \leftarrow \tilde{x}^{(M)} \). For the case of multiple passes, reset \( \tilde{x}^{(0)} \leftarrow \tilde{x}^{(M)} \) after each single pass.
2. Inner iterations: Update
   \[
   \begin{align*}
   \xi^{k+1}_m & = \xi^k_m + \omega Q^{(k)} - r \xi^k_m, \\
   \tilde{x}^{(m)} & = \tilde{x}^{(m-1)} + R^{-1} A^T_m (\xi^{k+1}_m - \xi^k_m) 
   \end{align*}
   \]
   where \( r = 0, \ldots, b - 1 \) and for each given \( r, m = 1, \ldots, M \). Running (12) and (13) through all \( M \) subsets once constitutes one pass.

Complete outer iteration:
Inexact solutions for \( \xi \) and \( x \) in the \((k+1)\)th iteration, after \( b \) passes, are: \( \xi^{(k+1)} \leftarrow \xi^{(k)} \) and \( x^{(k+1)} \leftarrow \tilde{x}^{(M)} \).

Two specific choices for \( Q_m \) are: (i) \( Q^{(k)}_m = (A_m R^{-1} A^T_m + V^{(k)}_m) \) and (ii) \( Q^{(k)}_m \) given by the diagonal of \( (A_m R^{-1} A^T_m + V^{(k)}_m) \). We refer to the corresponding BFS algorithms as BFS-SOR and BFSD respectively. Although we are uncertain of the general effects on convergence, we define a positively constrained BFS as:
\[
    x_j^{*,(k)} = \max(\epsilon, x_j^{(k)}). \tag{14}
\]
Constrained BFS substitutes \( \mu = Ax^{*,(k)} \) as diagonal of \( V^{(k)} \) for next iteration. This algorithm does not preserve the relationship between \( x \) and \( \xi \), but it is therefore necessary to reapply (7) at the conclusion of each outer iteration.

Equation (13) requires \( R^{-1} \) to update \( \tilde{x} \). A broad class of roughness penalties is provided by banded Toeplitz matrices, for which \( R^{-1} x \) is readily obtained as a local convolution filter operating on the image \( x \). With general non-negative definite \( R \), pre-computation of \( R^{-1} \) is an approach to reduce computations. Alternatively, we can specify \( R^{-1} \) directly. By specifying the form of \( R^{-1} \) we provide the prior covariance structure of the image \( x \).

For BFS-SOR and BFSD, we recommend storing iteration independent matrices \( A_m R^{-1} A^T_m \) and \( R^{-1} A^T_m \) and computing the iteration dependent diagonal elements when required. To compute the iteration independent matrices, note the fact that each column of \( A^T_m \) can be viewed as an image of the same size as \( x \). Hence \( A_m R^{-1} A^T_m \) is computed by filtering each of these images, with filter corresponding to pre-multiplying by \( R^{-1} \), then projecting the results to the corresponding block \( m \) of camera pixels.

We can show that if \( \mu^{(k)} > 0 \) for all \( k \), then, under certain regularity conditions, the unconstrained BFS algorithm converges locally to the MPL solution when \( k \to \infty \).

4. SIMULATION RESULTS
We compare positively constrained BFS-SOR and BFSD with BSREM and relaxed OS-SPS algorithms.

The simulation used an elliptical phantom of size \( 64 \times 64 \) pixels. A SPECT geometry was used with 64 attenuated parallel beam projections uniformly spaced over \( 360^\circ \), each projection contained 64 measurements. Attenuation coefficients were 0.15/cm (water) within the body, except for within the two lungs, where coefficients were 0.375/cm.

The projection matrix \( A \) (with dimension \( 64^2 \times 64^2 \)) was pre-determined by the geometry of pixels, adjusted according to the attenuation coefficients. Poisson noise was added to expected projections \( \mu = Ax \) to form the observed measurements \( y \). The total counts in projections was 400,605.

MPL reconstruction with quadratic penalty function requires choice of the penalty matrix \( R \). We specify \( R^{-1} \) rather than \( R \). We set the diagonals of \( R^{-1} \) to one. Non-diagonal elements were set to \( 1/4 \) if pixels were first-order neighbours, to \( 1/9 \) if pixels were second-order neighbours, and to 0 for all other cases.

We studied the positively constrained BFS algorithms with penalty parameter \( h = 10^{-5} \). These choices of \( h \) and \( \hat{R} \) may not provide the optimal smoothing in reconstruction, but suffice to study the convergence properties of available algorithms.

BSREM and relaxed OS-SPS algorithms, which require specification of \( R \), were coded and implemented in MATLAB and results were compared with that obtained from positively constrained BFS-SOR and BFSD. For BSREM and relaxed OS-SPS, one full pass through all subsets represents one iteration.

We ran BFS-SOR and BFSD with \( b = 1 \) and \( b = 8 \). All algorithms were tested with \( M = 64 \) blocks.

We adopt a notation combining algorithm name and blocks number; for example, BSREM-64 refers to the BSREM algorithm with number of blocks \( M = 64 \).

The starting values for BFS-SOR and BFSD were \( x^{(0)} = 0_{b \times 1} \) and \( \xi^{(0)} = 0_{b \times 1} \), but for BSREM and relaxed OS-SPS we used \( x^{(0)} = 1_{b \times 1} \) as relaxed OS-SPS exhibited poor behaviour from the zero start.

The relaxation parameter \( \nu \) in BFS was selected empirically. The choices provide near best early rate of convergence as judged by the penalized log likelihood at 10 iterations. Following Ahn and Fessler [8], BSREM and relaxed OS-SPS relaxations were \( \alpha_k = \alpha_0/(\gamma k + 1) \), with \( \gamma = 0.01 \). The initial relaxation, \( \alpha_0 \), was determined empirically similar to BFS.

To study the progress towards the MPL, Figure 2 display the log posterior ratios of the considered algorithms against their iteration numbers. This ratio is defined as \( \log l(\hat{x}) - \log l(x^{(k)}) \), where \( \hat{x} \) is obtained from 300 iterations of 1-pass BFSD-64 for the MPL reconstruction.

Figure 1 shows MPL reconstructions with different algorithms.

Figure 2 provides the log posterior ratio plots. At this level of smoothness BFS-SOR and BFSD converged faster then BSREM and relaxed OS-SPS. For example, in 64 iterations BSREM reached a penalized likelihood achieved by 16
Fig. 1. MPL reconstructions with smoothing parameter $h = 10^{-5}$. It displays BFSD-64 ($b = 1$, row 1, iterations 8 16 32 64); BFSD-64 ($b = 8$, row 2, iterations 1 2 4 8); BSREM-64 (row 3) and relaxed OS-SPS-64 (row 4) at iterations 8 16 32 64.

Fig. 2. Convergence towards MPL solution. It displays plot of log posterior ratio of BFSD-64, BFS-SOR-64, relaxed OS-SPS-64 and BSREM-64.

5. CONCLUSIONS

We conclude that BFS-SOR and BFSD are efficient block-iterative algorithms for MPL (unconstrained) with quadratic penalty. Under certain conditions these algorithms converge to the unconstrained MPL solution. Positive constraints can be enforced in BFS-SOR and BFSD by resetting non-positive estimates. In this case, however, we are only able to demonstrate convergence empirically; the simulation shows they converge to the same solution as BSREM and relaxed OS-SPS, algorithms with proven convergence to the positive MPL solution.

The simulation shows 8-pass BFSD achieved an acceptable MPL reconstruction in very few iterations. A good strategy is to use a large number of sub-iterations (such as $b = 8$) within each iteration providing good asymptotic properties with only small numbers of iterations.

Our simulation study suggests that BFS algorithms achieve a faster rate of convergence than BSREM and relaxed OS-SPS when the smoothing parameter is non-negligible. The diminishing step sizes appear responsible for poor performance of BSREM and relaxed OS-SPS. This result remains preliminary, given limited simulations and the fact that heuristic choices for relaxation parameters were made.

6. REFERENCES